

# REVIEW ARTICLE

## Evaluation of Affymetrix Gene Chip sensitivity in rat hippocampal tissue using SAGE analysis\*

Simon J. Evans,<sup>1,†</sup> Nicole A. Datson,<sup>3,†</sup> Mohamed Kabbaj,<sup>1</sup> Robert C. Thompson,<sup>2</sup> Erno Vreugdenhil,<sup>3</sup> E. Ronald De Kloet,<sup>3</sup> Stanley J. Watson<sup>1</sup> and Huda Akil<sup>1</sup>

<sup>1</sup>Mental Health Research Institute, 205 Zina Pitcher Place, University of Michigan, Ann Arbor, MI 48109-0720, USA

<sup>2</sup>Department of Psychiatry, University of Michigan, Ann Arbor, MI, USA

<sup>3</sup>Division of Medical Pharmacology, Leiden/Amsterdam Center for Drug Research, Leiden University Medical Center, Leiden, The Netherlands

### Abstract

DNA microarrays are a powerful tool for monitoring thousands of transcript levels simultaneously. However, the use of DNA microarrays in studying the central nervous system faces several challenges. These include the detection of low-abundance transcripts in highly complex tissue as well as estimating relatively low-magnitude changes in transcript levels in response to experimental manipulation. Many transcripts important to brain function have low expression levels or are expressed in relatively few cells, making them difficult to detect in the complex background of brain tissue. The aim of the present study is to evaluate the sensitivity of Gene Chip detection of transcripts in brain by using results from serial analysis of gene expression (SAGE) studies. The results of this comparison indicate that Affymetrix Gene Chips, like SAGE, only reliably detect medium- to high-abundance transcripts and that detection of low-abundance transcripts, many of which have great relevance to biological function in brain, is inconsistent. Specifically, we estimate that Gene Chips reliably detect no more than 30% of the hippocampal transcriptome when using a gross hippocampal dissection as the source tissue. This report provides the first broad evaluation of Affymetrix Gene Chip sensitivity relevant to studying the brain.

### Introduction

DNA microarrays are used to measure the relative expression levels of thousands of transcripts to enable global analysis of biological systems and have led to important discoveries, described in several reviews, in the fields of cancer (Clarke *et al.*, 2001), pathology (Kato-Maeda *et al.*, 2001), antimicrobial drug discovery (Ivanov *et al.*, 2000) and neuroscience (Cao & Dulac, 2001). However, the limits of DNA microarray technology have not been well described and the use of DNA microarrays to study the brain faces significant challenges with regard to sensitivity, largely because of the high complexity of neuronal tissue.

Estimates have been made to predict the limit of sensitivity of DNA microarrays. Specifically, Affymetrix, who manufacture Gene Chips—a form of DNA microarrays, reports that Gene Chips can reliably detect cRNA species in a hybridization cocktail at  $\approx 2$  pM. Using a homogenous cell population as the source of RNA this would translate to  $\approx 1$  copy per cell. However, this has little value in determining the limits of detection of microarray technology in complex tissue, like brain structures, where there are a large number

of specialized cells with variant transcriptional profiles. Many transcripts that are important for brain function are expressed at low levels or in a relatively small number of cells. For example, gonadotropin-releasing hormone (GnRH), which is imperative to reproductive function, is only expressed in a few thousand cells (Yellon *et al.*, 1990) dispersed throughout the hypothalamus (King *et al.*, 1984). Analysis of hypothalamic RNA by Affymetrix U34A Gene Chips only detects this transcript  $\approx 50\%$  of the time when using a gross hypothalamic dissection as the source tissue (our unpublished data). This insensitivity presents significant problems in successfully monitoring expression level variations in response to experimental manipulation of many biologically relevant transcripts.

Techniques to reduce tissue complexity, such as microdissection or laser-capture microscopy (Emmert-Buck *et al.*, 1996), are likely to improve the detection of cell-specific transcripts by DNA microarrays. These types of focused studies may indeed be the most appropriate at times but in many cases specific cellular nuclei important to a given experimental paradigm will not have been identified or the study of too many specific neuroanatomical areas by microarray may not be feasible. Typically, successful DNA microarray studies in brain have focused on discrete structures followed by the use of anatomical techniques, such as *in situ* hybridization (Mirmics *et al.*, 2000; Sandberg *et al.*, 2000; Mody *et al.*, 2001; Zhao *et al.*, 2001; Zirlinger *et al.*, 2001).

Serial analysis of gene expression (SAGE) is a high-throughput sequencing-based technique used to estimate the relative expression levels of thousands of transcripts by sequencing concatamers of short

Correspondence: Dr Simon J. Evans, as above.

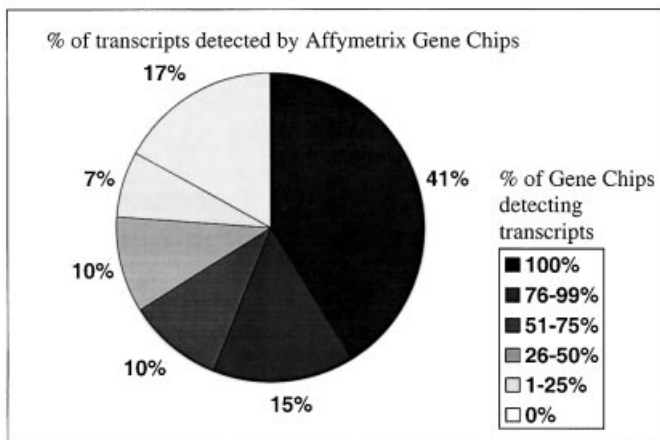
Email: evanssj@umich.edu

\*Presented at the workshop sponsored by the Serono Foundation on 'The Neuroendocrine–Behaviour Interface in the Post-Genome Era', 31 August – 1 September, Bristol, UK.

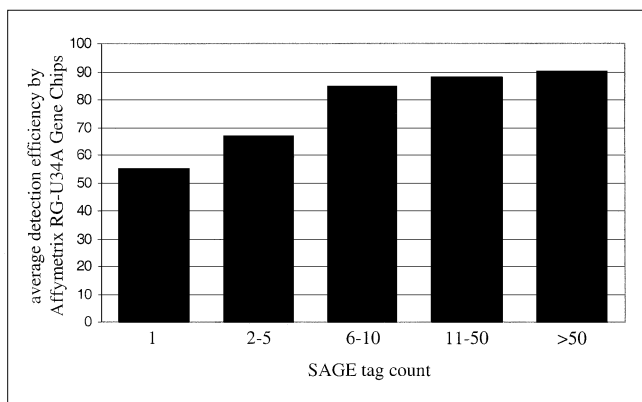
†S.J.E. and N.A.D. contributed equally to this work.

Received 5 April 2002, revised 17 May 2002, accepted 27 May 2002

Probability of Detection by Affymetrix Gene Chips of 1000 Transcripts in Common Between SAGE and Gene Chip Studies.



Average Gene Chip detection efficiencies of SAGE tag bins



Distribution of SAGE tag bins relative to detection by Gene Chips

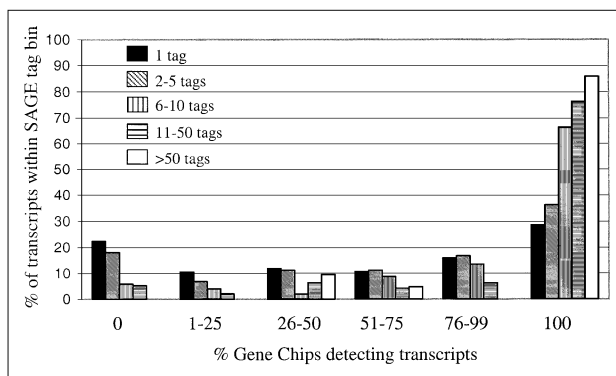


FIG. 1. Detection efficiency of Affymetrix RG-U34A Gene Chips in the detection of 1000 hippocampal transcripts, using expression level estimates obtained by SAGE. SAGE studies are compared with analysis using 43 independent Affymetrix RG-U34A Gene Chips. (a) The overall efficiency in detection of the 1000 transcripts by Gene Chips is shown independent of abundance predicted by SAGE. (b) The average efficiency in detection of transcripts by Gene Chips is plotted relative to their abundance predicted by SAGE. Each bar represents the percentage of the time each transcript within the given SAGE tag bin was detected across all 43 Gene Chips. (c) The distribution of transcripts of varying abundance predicted by SAGE is plotted relative to their detection by Gene Chips. The proportion of transcripts within each SAGE tag bin falling into different degrees of percentage detection by Gene Chips is shown.

sequence tags derived from a biological sample (Velculescu *et al.*, 1995). The relative abundance of transcripts in the sample can be estimated because it is proportional to the representation of the transcripts by sequence tags in the SAGE data. This technique has previously been used by Datson *et al.* (2001) to analyse the transcriptome of the rat hippocampus, which is a complex heterogeneous tissue composed of several morphologically distinct neuronal and non-neuronal cell types. In this study, in which a total of 76 790 SAGE tags were analysed, the number of low-abundance transcripts was shown to be far greater than the number of high-abundance transcripts. In fact, transcripts of low abundance, detectable with 1–5 SAGE tags each in this study, represented 80% of the total number of unique tags. Transcripts of intermediate abundance, with 6–50 tags, represented 19% of the unique tags, while high-abundance transcripts with >50 tags represented only 1% of total unique tags detected. Furthermore, annotation of transcripts in this study, to reveal distribution of functional gene families with regard to expression level, found that most genes contributing to functional specialization of neuronal tissue are of relatively low abundance. For example, neurotransmitter receptors, ion channels, synaptic components, signal transduction machinery and transcription factors were generally represented by few SAGE tags.

In the current study we have evaluated the efficiency of detection of hippocampal transcripts by Affymetrix RG-U34A Gene Chips using expression level estimates obtained by SAGE. Using this analysis we were able to estimate the percentage of transcripts being reliably detected by Gene Chips across various expression levels and functional classes. This report provides the first broad analysis of the sensitivity of DNA microarray technology in the complex background of neuronal tissue.

## Methods

### Animals and tissue preparation

Male Sprague-Dawley rats from Charles River (Wilmington, MA, USA), weighing 250–300 g, were used in this study. They were housed three per cage in 43 × 21.5 × 25.5 cm plexiglass cages, and kept on a 12-h light–dark cycle (lights off at 07.00 h). Food and water were available *ad libitum*. All experiments were conducted in accordance with the guidelines of the animal ethics committee at the University of Michigan following the Guide for the Care and Use of Laboratory Animals (National Research Council 1996). After 2 weeks of habituation to the housing conditions, the rats were killed by decapitation between 11.00 and 13.00 h. The brains were immediately removed, and hippocampi rapidly dissected and frozen. Tissues were stored at –80 °C until total RNA extraction.

### Gene Chip studies

In total, 43 Affymetrix RG-U34A Gene Chips were hybridized with cRNA derived from individual hippocampi from Sprague-Dawley rats so that each Gene Chip represented the hippocampus of one rat. Total RNA was extracted separately from individual hippocampi using Trizol Reagent (Invitrogen, Carlsbad, CA, USA) per manufacturers instructions followed by a cleanup step using RNeasy RNA purification columns (Qiagen, Valencia, CA, USA). Using 10 µg total RNA as determined by absorbance at 260 nm, first and second strand synthesis was performed per instructions in the Affymetrix Expression Analysis Technical Manual, ver3 (Affymetrix, Santa Clara, CA, USA). cRNA was synthesized using the Bioarray High Yield RNA Transcript Labeling Kit (Enzo, New York, NY, USA) and prepared for hybridization with RG-U34A Gene Chips per

Affymetrix instructions. Gene Chips were hybridized for 18 h at 42 °C, washed and stained on an Affymetrix fluidics station using the standard EukGEws2v3 protocol and scanned with a Gene Array Scanner. Raw image data were quantified with Affymetrix MAS 5.0 using all default threshold settings.

### SAGE studies

SAGE data was derived from the results of the study by Datson *et al.* using hippocampi from male Wistar rats (Datson *et al.*, 2001). Unique accession numbers from SAGE data were merged with Affymetrix Gene Chip probe set identifiers using ProbeMatchDB (Wang *et al.*, 2002).

## Results

The results described below are derived from 1000 transcripts that could be evaluated because they were unambiguous and common to the SAGE and Gene Chip data sets. These are all of the transcripts with a SAGE tag that was uniquely identified by a single GenBank accession number and was mapped to a single UniGene cluster (build no. 96) that was represented on the RG-U34A array. This set of 1000 transcripts had a distribution through the various SAGE tag bins similar to those described in the original SAGE study (Datson *et al.*, 2001) and is therefore assumed to be representative for the purpose of the current study.

Figure 1a shows the overall detection efficiencies by Affymetrix Gene Chips of the 1000 hippocampal transcripts. This figure illustrates that Affymetrix Gene Chips reliably detect 41% and are unable to detect 17% of these transcripts across all 43 Gene Chips used in this study. The remaining 42% of transcripts are detected somewhere between 0 and 100% of the time, as further detailed in the figure. Detection is defined as receiving a 'present' or 'marginal' call from the Affymetrix MAS 5.0 software using the default settings, which would identify any transcripts with a detection *P*-value of  $\leq 0.06$ . The average detection *P*-values (across all 43 Gene Chips) for the various bins detailed in the figure were 0.548, 0.285, 0.172, 0.070, 0.022 and 0.002 for the detection bins of 0, 1–25, 26–50, 51–75, 76–99 and 100%, respectively.

Figure 1b shows the detection efficiency by Gene Chips of the 1000 unique transcripts relative to their abundance bins as described by SAGE. These results show that the low-abundance transcripts, represented by a single SAGE tag, were detected an average of 55% of the time and that high-abundance transcripts, represented by >50 SAGE tags, were detected an average of 90% of the time on Affymetrix U34A arrays. Transcripts of intermediate abundance with 2–5, 6–10 and 11–50 SAGE tags were detected an average of 65, 85 and 88% of the time, respectively. It should be noted that transcripts with a SAGE tag count of at least 5 in this study are considered reliably detected by SAGE.

Figure 1c expands the SAGE tag abundance bins into their relative distributions across detection efficiencies by Gene Chips. This figure shows what percentage of transcripts from each SAGE tag bin was reliably detected (by 100% of the Gene Chips), undetected (by 0% of the Gene Chips), or detected unreliably (by between 0% and 100% of the Gene Chips). In general, a higher proportion of tags were reliably detected than were undetected or unreliably detected at all expression levels analysed. For transcripts of low abundance, represented by 1–5 SAGE tags, the undetected category was the second largest whereas, for the intermediate-abundance transcripts, represented by 6–50 SAGE tags, the intermediate detection efficiency categories were the second largest. Transcripts of high abundance, with >50 SAGE tags,

were never completely undetected by Gene Chips. To summarize, the detection efficiency of Gene Chips, in general, increased with the SAGE tag count. However, there were a high percentage of transcripts reliably detected by Gene Chips at all expression levels estimated by SAGE.

The glutamate system is known to play a key role in hippocampal function (Kullmann & Asztely, 1998; Bortolotto *et al.*, 1999), which is exemplified by the fact that glutamate receptors were the most frequently observed class of neurotransmitter receptors in the SAGE study (Datson *et al.*, 2001). To illustrate how an entire system is detected in the context of these broad analysis tools, we focus on the known components of the glutamate system that are found in both the SAGE and Gene Chip data sets. Table 1 highlights these transcripts and reports both their SAGE tag count and their efficiency of detection by Gene Chips. This table shows that increased SAGE tag count generally predicts increased Gene Chip detection efficiency within this system, as is the case for transcripts in general. Examination of the transcripts listed in the table show that many receptors are expressed at low levels and are outside the range of reliable detection for both techniques, whereas the transporter and synthesis machineries are expressed at higher levels and more reliably detected. Although the detection by SAGE and Gene Chips correlates quite well, there are some differences. For example, metabotropic glutamate receptor 3, kainate receptor subunit ka2 and glutamate transporter GluT are reliably detected by Gene Chips but not reliably detected by SAGE (< 5 tags) whereas NMDAR1 is reliably detected by SAGE (> 5 tags) but not reliably detected by Gene Chips.

## Discussion

In this study we evaluated the sensitivity of Affymetrix Gene Chips in a complex neuronal tissue, the hippocampus. These studies show that relatively high-abundance transcripts are reliably detected by Gene Chips but that a significant percentage of low-abundance transcripts are undetected or unreliably detected. In addition, a vast majority of expressed transcripts in the hippocampus fall into the low-abundance category. This work is important because DNA microarray studies are currently being used to provide a global view of the transcriptome and the current study estimates the degree to which this approach is successful in brain tissue.

Although there is a strong correlation between the expression level of a transcript predicted by SAGE and the detectability of the transcript by Affymetrix Gene Chips the relationship is not strict. The data show that Gene Chip probes complementary to transcripts expressed at similar estimated expression levels do not necessarily perform equally. Gene Chips may under-perform due to probe design issues, such as: distance of the target sequence from the poly-A tail; secondary structures within the target sequence; and cross-reactivity of the probe with other transcripts, each of which may influence detection. On the other hand, SAGE may under-perform because specific transcripts may be missed due to the absence of a recognition site for the anchoring enzyme or GC-content bias (Margulies *et al.*, 2001). In addition, incorrect tag counts can arise from incomplete digestion or alternative polyadenylation, giving rise to multiple tags derived from a single transcript. Also, sequencing errors, especially in the population of tags encountered only once, can contribute to the differences between SAGE and Gene Chip data.

Besides technical explanations, a biological explanation for observed differences between Gene Chip and SAGE data might be that male Wistar rats were used in the SAGE study and male Sprague-

TABLE 1. Expression level estimates by SAGE and detection efficiency by gene chips of hippocampal glutamate system components

Accession no.	SAGE tag count	Gene chip detection (%)	Description
M91652	38	100	Glutamine synthetase (gluA)
M38061	19	100	Glutamate receptor (GluR-B)
D63772	9	100	Neuronal high-affinity glutamate transporter
M92076	1	100	Metabotropic glutamate receptor 3
Z11581	2	98	Kainate receptor subunit (ka2)
U15098	1	98	GluT and GluT-R glutamate transporter
X63744	2	74	Glutamate/aspartate transporter
D87839	2	63	Beta-alanine oxoglutarate aminotransferase
U11418	6	53	NMDAR1 glutamate receptor subunit
Z11548	2	49	Glutamate receptor subunit (GluR6), kainate subtype
X54656	1	47	Glutamate receptor subunit (GluR-K3)
M83561	1	44	Glutamate receptor subunit 5-2 (GluR5-2), kainate subtype
U08260	2	28	NMDAR2D glutamate receptor subunit
X17184	4	21	Glutamate receptor, AMPA subtype, GluR1
U08255	2	14	Glutamate receptor delta-1 subunit
D13212	1	12	NMDAR2C glutamate receptor subunit
D13213	2	0	NMDAR2D-1 glutamate receptor subunit
AF027331	1	0	Kainate receptor GluR7b

The table highlights the components of the glutamate system that were uniquely detected by SAGE and represented on Affymetrix RG-U34A gene chips. The SAGE tag count and the detection efficiency by gene chips is reported.

Dawley rats were used in the Gene Chip studies. Although this may have a small impact on the conclusions drawn below, it is unlikely that it will be significant because these two strains of rats are closely related, as is evident by the overall good correlation in this study between the SAGE and Gene Chip data sets. Supporting an assumption of minimal strain differences, a study using Gene Chips to examine brain expression profiles of two different strains of mice found < 1% of the genes investigated to be differentially expressed in the hippocampus (Sandberg *et al.*, 2000).

The SAGE data from rat hippocampus (Datson *et al.*, 2001) utilized by the current study is probably not a complete representation of the hippocampal transcriptome. In fact, extrapolating the raw data of unique tags identified vs. tags analysed predicts that there are  $\approx 66\,000$  unique sequence tags ( $R^2 = 0.998$ ) expressed by hippocampus (data not shown). By these calculations the SAGE study under discussion, which revealed 28 803 unique tags, detected  $\approx 41\%$  of the rat hippocampal transcriptome. Thus, calculations relating to the percentage of transcripts detected by Affymetrix Gene Chips as reported in the results section are likely to be overestimates when considering the entire hippocampal transcriptome, because it is probable that the remaining transcripts undetected in the SAGE analysis would be skewed toward the low-abundance bins. Following this logic, it is likely that a large percentage of the transcripts that remain undetected by the SAGE study would also go undetected by Gene Chips. If we assume that the transcripts undetected by SAGE would be detected by Gene Chips no better than those transcripts represented by a single SAGE tag, we can estimate that Affymetrix Gene Chips reliably detect no more than 30% and fail to detect at least 30% of hippocampal-expressed transcripts using a gross hippocampal dissection as the source tissue. This leaves  $\approx 40\%$  of hippocampal transcripts falling into the range of unreliable detection.

This study suggests that Affymetrix DNA microarray technology applied to neuroscience can only reliably reveal regulation of mRNA species of medium to high abundance. This shortcoming applies to SAGE as well, given that the overall sensitivity of the two technologies in this study appeared similar.

The consequences of these limitations will obviously depend on the system under investigation. For example, receptors are generally expressed at lower levels than neurotransmitter/neuropeptide pro-

cessing machinery, as exemplified by the glutamate system in this report. Although there will be many exceptions to this, the consequences are that receptor regulation may be more difficult to detect by DNA microarrays than regulation of the more abundant synthesis or signal transduction machinery. However, a small number of insights into system regulation generated by DNA microarrays can lead to further discoveries within specific pathways by classical studies, such as *in situ* hybridization or immunocytochemistry.

To conclude, current DNA microarray technology is not likely to reveal regulation of low-abundance transcripts, which represent a large percentage of the transcriptome in complex tissue. However, provided that the limits of Gene Chips are understood, the technology does provide a powerful, state-of-the-art tool to analyse brain transcriptional profiles and can successfully lead to important discoveries by more focused studies of specific systems using other techniques.

## Acknowledgements

SAGE experiments were supported by the Netherlands Organization for Scientific Research, Grants 903-42-085 and 925-01-008. Gene chip experiments were supported by the Nancy Pritzker Depression Research Network and the National Institute of Mental Health, Program Project Grant 5 P01 MH42251.

## Abbreviation

SAGE, serial analysis of gene expression.

## References

- Bortolotto, Z.A., Fitzjohn, S.M. & Collingridge, G.L. (1999) Roles of metabotropic glutamate receptors in LTP and LTD in the hippocampus. *Curr. Opin. Neurobiol.*, **9**, 299–304.
- Cao, Y. & Dulac, C. (2001) Profiling brain transcription: neurons learn a lesson from yeast. *Curr. Opin. Neurobiol.*, **11**, 615–620.
- Clarke, P.A., te Poele, R., Wooster, R. & Workman, P. (2001) Gene expression microarray analysis in cancer biology, pharmacology, and drug development: progress and potential. *Biochem. Pharmacol.*, **62**, 1311–1336.
- Datson, N.A., van der Perk, J., de Kloet, E.R. & Vreugdenhil, E. (2001)

- Expression profile of 30,000 genes in rat hippocampus using SAGE. *Hippocampus*, **11**, 430–444.
- Emmert-Buck, M.R., Bonner, R.F., Smith, P.D., Chuaqui, R.F., Zhuang, Z., Goldstein, S.R., Weiss, R.A. & Liotta, L.A. (1996) Laser capture microdissection. *Science*, **274**, 998–1001.
- Ivanov, I., Schaab, C., Planitzer, S., Teichmann, U., Machl, A., Thöml, S., Meier-Ewert, S., Seizinger, B. & Loferer, H. (2000) DNA microarray technology and antimicrobial drug discovery. *Pharmacogenomics*, **1**, 169–178.
- Kato-Maeda, M., Gao, Q. & Small, P.M. (2001) Microarray analysis of pathogens and their interaction with hosts. *Cell Microbiol.*, **3**, 713–719.
- King, J.C., Anthony, E.L., Gustafson, A.W. & Damassa, D.A. (1984) Luteinizing hormone-releasing hormone (LH-RH) cells and their projections in the forebrain of the bat *Myotis lucifugus lucifugus*. *Brain Res.*, **298**, 289–301.
- Kullmann, D.M. & Asztely, F. (1998) Extrasynaptic glutamate spillover in the hippocampus: evidence and implications. *Trends Neurosci.*, **21**, 8–14.
- Margulies, E.H., Kardia, S.L. & Innis, J.W. (2001) Identification and prevention of a GC content bias in SAGE libraries. *Nucl. Acids Res.*, **29**, E60–E60.
- Mirnics, K., Middleton, F.A., Marquez, A., Lewis, D.A. & Levitt, P. (2000) Molecular characterization of schizophrenia viewed by microarray analysis of gene expression in prefrontal cortex. *Neuron*, **28**, 53–67.
- Mody, M., Cao, Y., Cui, Z., Tay, K.Y., Shyong, A., Shimizu, E., Pham, K., Schultz, P., Welsh, D. & Tsien, J.Z. (2001) Genome-wide gene expression profiles of the developing mouse hippocampus. *Proc. Natl Acad. Sci. USA*, **98**, 8862–8867.
- Sandberg, R., Yasuda, R., Pankratz, D.G., Carter, T.A., Del Rio, J.A., Wodicka, L., Mayford, M., Lockhart, D.J. & Barlow, C. (2000) Regional and strain-specific gene expression mapping in the adult mouse brain. *Proc. Natl Acad. Sci. USA*, **97**, 11038–11043.
- Velculescu, V.E., Zhang, L., Vogelstein, B. & Kinzler, K.W. (1995) Serial analysis of gene expression. *Science*, **270**, 484–487.
- Wang, P., Ding, F., Chiang, H., Thompson, R.C., Watson, S.J. & Meng, F. (2002) ProbeMatchDB—a web database for finding equivalent probes across microarray platforms and species. *Bioinformatics*, **18**, 488–489.
- Yellon, S.M., Lehman, M.N. & Newman, S.W. (1990) The gonadotropin-releasing hormone neuronal system of the male Djungarian hamster: distribution from the olfactory tubercle to the medial basal hypothalamus. *Neuroendocrinology*, **51**, 219–225.
- Zhao, X., Lein, E.S., He, A., Smith, S.C., Aston, C. & Gage, F.H. (2001) Transcriptional profiling reveals strict boundaries between hippocampal subregions. *J. Comp. Neurol.*, **441**, 187–196.
- Zirlinger, M., Kreiman, G. & Anderson, D.J. (2001) Amygdala-enriched genes identified by microarray technology are restricted to specific amygdaloid subnuclei. *Proc. Natl Acad. Sci. USA*, **98**, 5270–5275.